Review and Discussion: Examining the User Experience Implications of Active Variant TLDs, Draft Final Report January 2013[1]

John C Klensin, 2013-02-27

General Observations

This report is a nice piece of work, reasonably easy to follow and clear about the issues it discusses. The authors are commended for managing to create an exception to recent patterns by keeping it under 100, indeed under 75, pages. The vast majority of the analyses and recommendations are sufficiently clear that a reviewer can understand and evaluate them in context, rather than having to puzzle out what was intended. Partially for those reasons, this version is also a very significant improvement over the earlier draft from this component of the Variant Information Project. Nonetheless, the report seems, in several places, to miss key points and to have difficulties with scope, concepts, and general aspects of the systems of which IDNs and variant IDNs are a part.

This review consists of two parts. In the first one, I address major issues with the report, issues significant enough to call its validity and utility into question. Some of them identify key questions that the report does not address. In the second, I make section-level comments about parts of the analysis and recommendations of the report and, in many cases, elaborate on my comments about the major issues in the specific context of the text of the report.

Major Issues

1. Scope: Internationalized Domain Names, Character Repertoire, and Variants

Like the LGR report, this report is not limited to variant issues but instead addresses a large number of issues with IDNs generally, including the character repertoires that should be allowed at the top and lower levels, and even some topics that are applicable to more traditional all-ASCII domain names. From my point of view, that expansion of scope beyond what the community probably anticipated is actually very desirable because it permits looking at the systems involved rather than isolated corners of them that cannot be put together with the whole. As noted below, there are some areas in which I do not believe the process went far enough even with the scope expansion. However, it is possible, indeed likely, that some groups of stakeholders in the broader community have believed that they can ignore this work because it is strictly about "IDN variants" and they have concluded that they need not be concerned with that topic. It would be therefore be desirable for ICANN to ensure that there is broad understanding

---

of the topics and recommendations addressed by this report so they can be reviewed in a broader and more appropriate context.

Equally important, there are a number of recommendations in the report for specific actions or activities which, if they are needed at all, would be needed for IDNs even if no variants were ever allocated or activated. It would have been much more helpful if the report identified those so that the reader could understand what, as its title suggests, the actual implications and effects of active variant TLDs are given IDN TLDs (or new TLDs generally, see item 3 below), rather than lumping everything the writers identified as useful together independent of specific cause or motivation.

2. The Linguistic Assumption, the DNS, and User Interfaces

The report makes the very fundamental assumption that the best (or only) way to deal with the "linguistic" and user expectation issues it identifies is by some Domain Name System mechanism identified with the term "variant". That is not the case and, interestingly, the report provides foundations for some of the other options. By assuming that "variants" are the preferred or only solution to many of the expectations outlined in the report, the report tends to lead the ICANN community into a syndrome that is often described in such metaphorical terms as "if all you have is a hammer, everything looks like a nail". However, especially if ICANN were to engage in some of the user education and applications support activities recommended in the report, there may be other options that pose less risk to the DNS and to Internet stability and security than trying to use DNS mechanisms to address all such issues.

3. Types of Variants, Character and Otherwise

The report correctly notes that relationship among labels that might reasonably be considered to be "variants" can arise from many sources of which the character-based variants that are the subject of label generation rule work are only one, one that is very important for some scripts and not for others. It then largely ignores those other cases, focusing almost exclusively on LGR-produced variants. In designing variant mechanisms, it is important to note that, while they weren't called that, we've had multiple labels in the same or different domains representing the same general (or closely-related) context and under the same management almost since the dawn of the DNS. As handy illustrative examples, I note the very old registrations and delegations of "digital.com" and "dec.com" and, in different TLDs, the three domains "toronto.edu", "utoronto.ca", and "utoronto.com". More recently, we've had "sony.com", "sony.jp", and I assume by now the corresponding IDN. In the all-ASCII space, registrants who have been concerned about difficulties or attacks using synonyms or spelling variations have, in many cases, simply registered the alternate strings, either "parking" them (when allowed by various rules) or delegating them for highly controlled management. Some of these cases have arisen from acquisition. For example, National Semiconductor had and used "national.com" for many years. When Texas Instruments acquired that company, many of the relevant names were redirected (either in applications or by changes to targets of DNS entries) to their equivalents in "ti.com" which, incidentally, is managed as synonymous with "texasinstruments.com". Some registrars have even built business models around identifying such possible conflicts, pointing out how expensive and aggravating it can be to challenge a confusing or trademark-violating registration by another party, and encouraging defensive registrations as an alternative.

It seems to me that there are two important lessons from that history, neither of which is on the radar of the report. First, we have survived reasonably well for more than two decades without specific mechanisms for bundling labels, keeping special databases, having specialized "variant" provisioning tools, designing a user experience around the ability of users to predict which alternate names will be delegated, and so on. That has been managed by allowing registrants to apply for (and pay for) the names they want, defending those names by preemptive registrations, challenge and objection mechanisms, and dispute resolution procedures or legal action when needed. The report *assumes* that special mechanisms for variations are required; it does not make a persuasive case that the systems and mechanisms it recommends in each of its many categories will add enough marginal value to security, stability, or usability to justify the considerable costs in resources and changed procedures that those recommendations imply. My own guess is that the answer may turn out to be different for different recommendations, with some being justified and others not. Differentiating among them requires a more nuanced examination than the report attempts about what problems are being solved and which ones would cease to exist or become minimal under different scenarios.

Second, the arguments in the report (some only implicit) for special handling of (label) variants suggest that, if bundling techniques are necessary for IDN variants or character-based ones, then they are probably desirable, retroactively, for some or all of the historical registrations of related names in the same zone. The fact that we did not know for several decades that we had a problem that needed a solution does not force a conclusion that there was actually no such problem. However, if that is correct and there is as much reason to manage "ti.com", "texasinstruments.com", and perhaps "national.com" together as there is to manage, e.g., a Simplified Chinese and Traditional Chinese pair of labels together, then much of the discussion in the report is not tied to IDNs at all. In addition, examples of related names in separate zones — inevitable if variants are allocated in the root— should cause ICANN to think about whether cross-zone bundling (and perhaps blocking) is appropriate.

4. DNS Labels and Fully-Qualified Domain Names (FQDNs)

While DNS registration and similar activities operate on labels, many uses of the DNS, including URLs, email addresses, and most DNS-based security identifiers, depend on fully-qualified domain names. The report does not make the distinction clearly even though it is possible to deduce what is intended from the context of most sections. For the purposes of the report, the difference is important for at least one major context: with most character variant scenarios that involve actual delegation of more than one variant in a bundle, the number of variant labels within a single zone will be bounded by the bundle generated by the LGR or equivalent mechanism and whatever limits are set on the maximum number of labels to be delegated, but the number of FQDNs that must be managed —and, to the extent to which the end user actually needs to understand variant relationships, understood by that user— is multiplicative with the number of labels in the FQDN that need variant treatment. In other words, if, e.g., the Arabic or Devanagari cases called for three delegated variants (including the primary one) for a typical label, we would have for a two-level name:

> SLD1.TLD1, SLD1.TLD2, SLD1.TLD3, SLD2.TLD1, SLD2.TLD2, SLD2.TLD3, SLD3.TLD1, SLD3.TLD2, and SLD3.TLD3

for a total of nine FQDNs (nine is at the very upper limit of what the psychological literature tells us a typical human can keep track of). For a more typical three-level name, we would have 27 FQDNs, and so forth, with the situation becoming even worse if more than three variants were permitted to be delegated.

The dichotomy between Simplified Chinese (SC) and Traditional Chinese (TC) (a situation that does not exist in any of the other scripts studied in the VIP activity) can help considerably if registrants and delegations are confined to all-Simplified and all-Traditional trees with no mixing, creating a case parallel to the above with only the two SC-SLD.SC-TLD and TC-SLD.TC-TLD FQDNs. But, if that discipline is not enforced and, instead, the SLD registrant is permitted to register an SC label, a TC label, and a third label of their choice in each TLD as the report suggests is now the practice, then we return to a multiplicative situation similar to the one above with

> SC-SLD.SC-TLD, TC-SLD.SC-TLD, Arbitrary-Mixture-SLD.SC-TLD, SC-SLD.TC-TLD, TC-SLD.TC-TLD, and Arbitrary-Mixture-SLD.TC-TLD,

assuming that the "mixed" case is not allowed in the root. For all of the reasons discussed in the report, having to manage the relationships among a large number of FQDNs is, at best, an administrative nightmare. It also raises other issues, discussed elsewhere in this review.

5. The Variant Contradiction

As the report at least hints in several places, the key to a positive experience for the end user is predictability (which, in turn, depends on consistency): the user should know what is likely to work and should not be astonished when it either does or does not. Several of the recommendations about user education and educational materials appear to be aligned with that principle. While it would make few people happy, especially those who have become convinced by a long serious of discussions and task groups that they merely need to pronounce the word "variant" in the appropriate tone for all sorts of magical and desired things to happen, it would be very easy to explain to users that names are, in general, represented in only one way and that they should get used to it. Of course, that would directly contradict a basic premise of this study that there are at least some cases in which delegation of more than one variant from a bundle is sufficiently important to be allowed. If they are allowed, then the next simple story to tell users from which they can make predictions is that all of the orthographic variations they can reasonably think of will work. However, if a limit is placed on the number of variants that can be delegated —something the report strongly recommends— then the user will not be able to predict what will work and what will not, at least without mastering potentially complex rules that may differ by script and domain, and will not be able to predict at all if which variants are activated is at the discretion of the registrant. While the report skirts the issue, that is the worst possible case from a user experience perspective, taking the form of "sometimes the alternate string I enter will work, sometimes it won't, and there is no way to predict at all". That case is even worse than "there is only one form of the name and you need to get used to it even if it is hard to type".

Consequently, from that particular user experience perspective, either all plausible variants in a bundle should be placed in the DNS or at most one of them should be. Anything else leads to a bad experience. The report argues convincingly that delegating all plausible variants is a bad

idea from both user experience and security, stability, and other perspectives, so we have a contradiction.

It is not the only contradiction. The report argues, for what I consider good and convincing reasons, that variants should not be allocated, much less delegated, unless they are really important for a good user experience. Again using language that is more blunt than the report, that means that a second label from a bundle (or even more of them) should be allocated if and only if the applicant makes a convincing case that failure to do so will significantly interfere with the usability (or security, stability, etc.) of the primary name. However, if failure to allocate the (additional) variant has that negative an impact on the primary name, then the primary name should not be delegated unless the additional variant(s) can be allocated and delegated at the same time: if the primary name is adequately usable without the additional variants, then the variants are not necessary. Turning that into a policy guideline would be fully consistent with the report as I read it. However, it implies that any application that is now in the queue that requests that a variant be delegated is in a difficult position: either the variant is unimportant, in which case it should not be delegated (ever) or the variant is important, in which case the requested primary label should be withheld until ICANN is ready to consider allocating and delegating variants. One could go ahead and delegate the primary string on an exception basis, but the contradiction lies in the application itself (and willingness to accept one delegation without the other(s)), not in the system. And every exception or special case is going to make the overall user experience less predictable.

6. End Users and Variant "Handling"

Several places in the report suggest that end users will need to understand, be educated about, and adapt to variants and that the applications that support those users will need to be modified to be variant-sensitive. From one perspective (and to be a bit more blunt than the report), delegated variants are a DNS hack to compensate for end user inability to guess exactly how a string is written or for difficulties in entering some forms but not others. If the users or their application software have to understand that variants are special, how a bundle is composed, etc., then the hack fails and the variant mechanism doesn't accomplish anything very useful.

Conversely, assume that the introduction of top-level variants really does require ICANN (and TLD registries and registrars) to mount massive end-user educational efforts, that ICANN (for the root) and TLD registries (for those zones) keep "bundle" databases available for real-time access and searchable by any of the names in the bundle, and that end-user interface software systems be modified to be variant-aware. If those conditions hold, we should be moving a further step beyond "variants in the DNS are the answer, almost independent of the question". For example, if an application could query the bundle database at lookup time and use it to determine which member of the variant set should actually be used in a DNS query, then a very large fraction of the problems mentioned in the report would simply disappear: there would be no issues associated with equivalent names in the DNS because there would be only one name in the DNS and there would be no issues associated with the user not being able to access the right variant (e.g., because it was not delegated) because all of the variants in the bundle could be considered without adverse consequences for the DNS. Of course, that approach would not be trivial: a new database with performance, robustness, and security/integrity features at least equivalent to the DNS would need to be designed and deployed. But it would provide a far better and more predictable user experience than DNS-based variants, would isolate tasks in

ways that would make many things easier for registrants, registrars, and registries than what is recommended in the report, and might not take significant longer to deploy than the systems the report calls for (and far less time than trying to modify the DNS to support, e.g., a modified query model or even new required RR types).

7. Blocking, Withholding, and Other Options

This report claims to be agnostic about the decisions as to how a variant that is identified and made part of a bundle is assigned to a state (see Section 6.1.4 and the Integrated Issues Report (IIR)) but then proceeds to focus on activated and delegated variants as its title suggests. The decision to separate choices of state from the report has important consequences in terms of questions that are not raised and issues that are not addressed. In particular, suppose a decision were made that the disposition of all variant bundles would permit a maximum of one label to be activated and that other labels would be blocked (or permanently withheld). That would turn the variant picture into a proactive and low-overhead mechanism for keeping excessively similar names out of the DNS and would dramatically change much of the analysis in Section 5 of the report.

8. Characterization of the Position of the "Technical Community"

As someone who would almost certainly be considered part of the technical community, I believe the characterization of that community in the report, particularly in Section 1.4, is inaccurate. Many of us consider a high-quality and predictable user experience, one that is sensitive to a range of issues with data entry and character rendering and output, to be fully as important as the sort of narrow view of security and stability issues represented in the report. It is likely that, compared to other groups, a larger fraction of the technical community understands the inherent limitations of the design of the DNS and are consequently more hesitant about the possible consequences of trying to "trick" the DNS into doing things that lie outside or in contradiction to that design, but that is a separate issue. Although at least some of us see many issues with an expansive view of variants (some of which are reflected in this review), those concerns come from efforts to understand and balance the issues, not from the narrow focus the report suggests.

9. High-Level Conclusion and Recommendation

For me, each new report from the Variant Information Project activity seems to reinforce a tentative conclusion that I reached when the first-round team reports started to appear and have commented on earlier. This is all just too complicated and, as this report (even without the additional issues identified in this review) clearly shows, creates significant additional risks and costs without being able to offer the payoff of a predictable and consistent experience that will meet the needs of end users. Adjustment of expectations to match reality and considerable simplification appear to be in order. Most obviously, that might be to simplify and redefine the LGR activity to an effort to define a repertoire for the root and make general repertoire recommendations for SLDs (and perhaps below), to ask the team that produced the present report to redo it to focus exclusively on root IDN issues rather than IDN variant ones, and to deal with requests for related names in the root zone —especially those not associated with historical ccTLDs — on an exception basis with the applicant having to demonstrate to the community that the relationship is important enough (and will solve a specific problem that can be identified) to justify the various "challenges" and other difficulties outlined in this report.

Interestingly, while the above is much less complex (the ratio of one paragraph to the bulk of text in this report and the LGR one is somewhat indicative), the net result of this model compared to the one outlined in the LGR and User Experience report and recommendations might be only very slightly different.   The LGR report creates a process that will make it very difficult to get global agreement on labels that can be activated.   The recommendations of the present report set a very high threshold for actual activation of more than one variant in a bundle and justify those recommendations by identifying a wide range of risks and problems. So, net and with the understanding that other options might be possible, ICANN seems to have two options:

- Follow roughly the outline above with the understanding that it will result in very few variants being delegated (other than what would be the primary name in a bundle) although no one would be prevented from applying for additional names subject to the usual reviews and challenge procedures. While I believe it would be better to see if a simplified version of the LGR process (perhaps even skipping the generation panels for the scripts for which we have significant experience and confidence) can produce a satisfactory root repertoire and to get a revised and rethought version of this report as a set of guidelines about issue to be considered, ICANN could, in principle, solicit the required additional information and start considering the variant requests that are implicit in the applications now in the queue at its convenience.

- Follow the general model of the LGR report, including multiple panels and rules for generating and establishing the status of character variants; create mechanisms (so far undefined and undiscussed) for generating or otherwise establishing candidate variants that cannot be produced from character-based generation rules; and implement the many recommendations of the present report (perhaps informed by the comments in this review).   With the understanding that the two or three years (at least) those processes would be likely to take would increase the risks from the contradictions discussed above, expect to end up with few, if any, activated variants that would not be activated under the above model.

Comments on Specific Sections

These comments are less important to an overall understanding and application of the report than those above, but may be useful in formulating the final version of the report and policy decisions based on it. Some of them are fairly significant in that they point out problems with the reasoning or conclusions of the present report.

10. Section 1

It would be useful if the second paragraph further separated out how many of the 31 Fast Track IDN ccTLDs are associated with variants (either allocated or requested). The end of that paragraph says "...IDN TLDs will also be part of ICANN's new Generic Top-Level Domain (new gTLD) program". The future tense may no longer be appropriate there and information about the number of IDNs applied for and the number of variant requests would be helpful for those evaluating the consequences of this report and the contingent risk of allocating and delegating names now that might not conform to further LGR

specifications (resulting, as the report notes, in inconsistencies and a diminished quality of user experience) or fall into the "is an allocated variant necessary" trap discussed in item 5 above.

11. Section 1.1

The first sentence, "An IDN TLD label may have many variants." is obviously true but not particularly useful without a discussion of the different models and relationships that determine those variants. The report mentions LGR and a few other possibilities but basically glosses over this. A little more detail would be useful.

12. Section 1.2

The definition from the IIR is very general and consistent with the comments and concerns above about many types of variants and relationships. Then the way in which the second full paragraph is written appears to constrain this report to variant labels generated by the LGR process. Specifically, if the subordinate clause "not just a code point or a character," were dropped from the sentence, it would indicate that the only labels under consideration were LGR-generated. The rest of the report appears to be inconsistent as to whether that is the intent or whether the sentence is just badly written and variant labels generated or determined in other ways are included. I believe it would be a grave disservice to the community to limit the report to LGR-generated labels.

Paragraph 4 ("There are ongoing discussions…") is actually not true or should not be. As the report points out, having variant treatment be similar as one moves down the DNS tree is very important to a consistent and predictable user experience (how similar should be reasonably expected is another issue). Since DNAME-based approaches can support variants at only a single level of the tree, effectively supporting them for one label but preventing them for others in an FQDN, they are inconsistent with this principle. Absent new alias models or a new DNS resolution model (either of which would almost certainly require a next-generation DNS protocol and infrastructure), the *only* remaining possibility for activating a label is to delegate it using existing, normal, mechanisms. Encouraging people to fantasize about other solutions or approaches to be deployed in the near future is not in the best interest of either ICANN or the broader community.

13. Section 1.4

To elaborate slightly on item 8 above in this context, many of the issues that have led to discussions of variants, including those resulting from Unicode decisions to assign multiple code points to essentially identical glyphs and, in other cases, to treat rather different glyphs as representations of the same abstract character and consequently assign the same code point to them are technical issues identified by the technical community. Those issues were discussed at great length as part of the specification of IDNA (both versions) and various non-DNS string comparison models that may affect the use of domain names in practice. Some of the issues that the report refers to as "linguistic reasons" (a categorization to which many linguists would object) are, themselves, related to the relationships between computer representations of coding, languages, and strings of characters and actual writing systems and hence very much "technical".

Especially where user perceptions are concerned, script-based distinctions in the root are feasible for some scripts and not for others. The Unicode definitions of scripts and their boundaries might not be optimal for DNS purposes. In particular, as has already been observed in the ccTLD IDN Fast Track process, script-based distinctions may not be feasible for very short strings drawn from closely-related scripts that share many character glyph forms, such as Greek, Latin, and Cyrillic. If variants are a solution to either confusingly-similar strings or strings whose script or coding cannot be reliably known from a printed form, then those scripts are likely to call for cross-script variant labels. To say that "script distinction is (arguably) possible" in the root is to skim over a range of issues that could have profound implications for both the user experience and for the feasibility of the LGR process.

14. Section 2.1.1
The next-to-last bullet, "A registrant may activate up to five variants at a time" is ambiguous as to whether what is intended is "A registrant may have up to five variants active at the same time" or "A registrant may activate up to five variants in a single application but may later apply for additional variants without removing any of the original five."

15. Section 2.1.2
The fifth bullet contains "further restriction by only allows a label". Was "… but only allows…" intended?

16. Section 2.1.4
Given the concern about multiplicative explosions of FQDNs discussed in item 4 above and elsewhere, it would probably be helpful to note that the (Latin) four-character SLD string "cira" is responsible for a bundle with nine variants (2 forms of "c", times 3 forms of "i", times 1 form for "r", times three forms of "a"), that a longer string would increase this number, and that applying the same rule to a third-level label would multiply the effect (for example, the seemingly-trivial FQDN cira.cira.ca would involve a bundle with 81 FQDNs in it). If the second and third level registrants decided to activate any significant fraction of these, the implications for URLs, caching, certificates, and so on would be significant (as discussed in Section 5 of the report).

17. Section 2.1.5
Part of the analysis in this section has an unnoticed (or at least un-noted) interaction with the specifications of the ccTLD IDN Fast Track. Since that procedure prohibited applications for Latin script strings, even distinctively non-ASCII ones, it is not possible to make any assumptions about whether countries who allowed IDN SLDs under their original (ISO 3166-1 alpha-2) ccTLDs would now be transitioning those subdomains into the new IDN TLDs for which they were prevented from applying.

If the stated TWNIC requirement for the registrars to manage DNS hosting and operations is important, a discussion of what occurs at the third level —in particular, whether those domains can be delegated and, if so, how variants are expected to be managed at that level — is important. If, on the other hand, TWNIC has made a decision to effectively eliminate the use of the DNS as a distributed administrative hierarchy, then little can be learned from their experience that is useful to the rest of the community (perhaps so little that they should have been excluded from the report).

18. Section 2.2

In the next-to-last bullet, was that intended to be "IDN.IDN"? Are the authors of this report confident that notation will be understood by all readers?

19. Section 3.3

It is probably worth noting that, while users may well have the expectations described in the second paragraph, there is little or no way of enforcing that assumption, especially below the second level, and no way at all to guarantee it. Another aspect of the principles leading to a predictable user experience is that leading users to expect levels of security-style protections that are not actually present can create the proverbial false sense of security and cause significant security risks in practice.

20. Sections 3.4, 3.6, and 3.7

As is probably clear from other parts of this review, I believe that the closely-related principles of Predictability and Consistency are pervasive, interacting in important ways with many of the other principles stated here as well as with both challenges and recommendations. However, predictability is a function of experience and extrapolation from one environment to another, rather than being, e.g., an inherent characteristic of the human condition. People can be educated about what to expect and disabused of incorrect inferences. For example, in many countries, a horizontal tray device in a car that slides in and out (possibly under control of a motor or spring) is a cup holder. Those users who extrapolate from that knowledge and experience to superficially-similar horizontal tray devices on computers and assume those are also cup holders tend to adjust their expectations rather quickly and not make the mistake a second time. Although there are exceptions for special circumstances and for "special" users, most computer users have had sufficiently dramatic experiences of one sort or another to not expect accurate "do what I mean" behaviors in response to whatever instructions they give or actions they take.

With that in mind, strong statements about what users expect and expect to be supported are at least somewhat questionable. Do users expect British and American spellings to match or a Latin script string with diacritical markings to match a string with those stripped off? The most accurate answers are probably "sometimes" or "it depends on their previous training or experience in what they believe to be the same context". Those users who have observed that there is no universal agreement as to how the strings ABC, bcd, and CDE are correctly ordered in a sorting or "alphabetizing" operation may have different expectations about what matches and what does not than users who have been oblivious to that particular set of relationships.

Even perceptions of what is or is not easy to use may depend on experience and training. There has been a debate in user interface design at regular intervals for nearly a half-century as to whether systems should be designed so as to be easy and efficient to use for the first-time user or whether ease of use and convenience for the more experienced user (perhaps even after significant training) is more important. The question can, and has, prompted arguments of intensity often reserved for religious disputes and, like most such questions, does not have clear and universal answers. But it is perhaps worth noting that a user on whose behalf dozens or even hundreds of domain names are resolved every day is almost certain to develop intuition and expectations for how the DNS works rather than returning each day to more abstract ideas about

how their languages work and should be supported.   They may not like the consequences of those intuitions, but most will learn them and learn how to work with them.

So we should keep in mind that we actually have a dual role that we need to balance. Part of it is to design our systems to accommodate the expectations of the users — to provide them with predictable and consistent experiences— and the report seems to be designed around that part. But the other part is to help those users understand what they should expect and predict. We need to balance the two, not assume that the entire burden can or should be shifted entirely in one direction or the other.

21. Section 4.1
It is probably important to include the use of identifiers that use DNS elements — security credentials, certificates, email addresses as login IDs, and so on— as part of this list of functions. As the report hints elsewhere, the constraints and requirements for such identifiers are different from those for, e.g., interactive web browsing.

22. Section 4.2 – Registrants and Registrars
In this section, the discussion of trademarks, and elsewhere, the report assumes that all registrants and related actors will have intentions that are positive for the Internet and its end users. We know from experience that is not the case. Some registrants acquire domains with the intention of launching attacks, often attacks that rely on confusion about names for their effectiveness. Some registrars take the position that the registration fees from those potential registrants are as good as the registration fees from anyone else and that refusing to serve them would be discriminatory and possibly even in violation of their agreements with ICANN.   It is perhaps unfortunate, but an ICANN mechanism that identified related names but did not immediately reserve or withhold them would probably be of great value to this group of would-be registrants and the registrars who support them.

23. Section 4.2 — Registries
Given the state of the new gTLD process, this description should probably be expanded to include registries that expect to deal with IDNs, not just ones who are already doing so. I also note that I can find nothing in the AGB that requires anything but the very crudest similarity of registration policies (including IDN-related ones) across ICANN-contracted gTLDs. Indeed, the "sponsored" category from the second-round gTLD group rather explicitly anticipated very different registration policy models.

24. Section 5.1.6
Statements like "user may not be able to input all of the variants" are, with rare exceptions, probably false. It would be more accurate —and would put a better perspective on the issues— to say that users would not find it acceptably convenient or would need to learn special skills in order to do so. This distinction is particularly important given the recommendations for training and training materials and the comments in item 20 above.

25. Section 5.1.10
I don't understand this section and suspect that it may need reconsideration and rewriting. Users don't typically search for an FQDN, whether variants are present or not, at least unless they are very confused. A user might include strings that match some of the labels of an FQDN in a

search query, but no contemporary search engine I know of gives DNS labels priority over content-related clues. By contrast, to the extent to which end users rely on search engines rather than direct entry of URIs, variants and other techniques that make it easier to match user intentions with primary labels in the DNS become less relevant in all respects (as well as a disadvantage as Sections 5.1.10 and 5.1.11 point out).

26. Section 5.1.16 (new)
There are some very complex issues associated with storage of information in cookies and what hosts and systems can retrieve those cookies. Those issues become more complex if "equivalent" names are delegated or otherwise established in the DNS and deserve some discussion in Section 5.1 of the report.

27. Section 5.2.1 and elsewhere
This report should explain, or reference something that explains, why the notion of a "primary" name is important. Unless there are aliases involved, the DNS cannot know the difference. Identification of some labels as "primary" may be convenient for some management and provisioning models and databases, but the topic needs more explanation before one worries extensively about, e.g., redesignating the primary.

28. Section 5.2.2
It is worth mentioning, and probably exploring, the fact that the problems discussed in this section become far more challenging when the subdomains of SLDs (3LD and 4LD) are considered and where their labels may have variants of their own.   In the general case, the issues and complexity are multiplicative, not merely an extra layer.

29. Section 5.2.3
Either the statement "ASCII TLDs and IDN TLDs are not variants of one another" has completely confused me, or the word "variant" is used in that sentence with a different meaning than the very broad definition that appears to be used in Section 1.2 (discussed in items 1 and 12 above).   If the user believes that the two TLDs are closely associated or have the same "meaning" and intent, then either they are variants or the Predictability and Consistency principles are violated. ICANN could, of course, make a policy decision as to whether to manage them in whatever special way that variants are managed, but that is a separate issue.

30. Sections 5.2.7 and 5.2.8
Trademark protection is one side of a problem. The other side is the kinds of registrants discussed in item 22 above. Until and unless ICANN succeeds in adopting enforceable policies that make registrations that are intended to take advantage of confusion (or even referral commissions based on typographical errors or the like) invalid, reports like this one need to consider those cases as legitimate as the cases of those who want to protect their trademarks.

31. Section 5.3.1
The last paragraph contains a note that "tens of variants are possible".   Web configurations and the need for web servers to know their own names depend on FQDNs, not label variants. Should a multilabel FQDN, with variants for each label (level), be used, tens of

variant FQDNs are likely and hundreds of them are quite possible. An order of magnitude change in the scale of the problem is probably significant.

32. Section 5.3.2

IDNA2008 discourages the use of the term "Punycode" as a noun for precisely the reason the third paragraph of this section illustrates. The Punycode encoding of the first U-label is "mgbai9a5eva00b". The A-label is "xn--mgbai9a5eva00b". Introducing the Punycode terminology here does the reader no favors, especially when it is used incorrectly.

As another part of the configuration issue, it is worth noting that web servers usually need to know there own names and that mail servers are absolutely required to do so: some of the names or an abstract concept of "variants" is not sufficient.

33. Section 5.3.3

This section or a separate one should discuss the very important issue of certificate revocation, especially the need to do so on an emergency basis. The existing text examines the question of obtaining the certificates, which may be burdensome but is easy in principle — all of the needed names are known. But revocation of a cluster of certificates linked to a variant cluster requires being able to retrieve all of the names and credentials associated with that cluster and take action on them.

34. Section 5.3.5

As with other examples and subsections, this section is written in terms of an SLD.TLD relationship, but things become multiplicatively more complex when three or more levels are involved, all potentially with delegated variants.

35. Section 5.3.7

The second sentence says "As caching matches domain names, it may not work effectively…" First, caching matches URLs, not variants, with all of the problems of URL matching discussed elsewhere in the report. Second, unless the cache understands the entire set of relevant variant structures and fundamental changes are made to the standards for web caches, it is certain that it will not "work effectively"… there are no other possibilities.

The second paragraph is also too optimistic. The report, experience, and various other specifications anticipate both variant FQDNs that "point to" the same content and ones that "point to" related but different content (e.g., content localized to match properties of the particular FQDN used). Even if an algorithm could be developed to permit identifying variant FQDN sets and the URL matching rules used in caching changed to match, it is almost certain that these two different types of variant applications could not be distinguished from each other without very significant changes in the relevant architectures.

36. Section 5.3.9 (and, to some extent, 5.3.5)

While variant domains are, almost by definition, related, whether they are equivalent or not depends on some of the factors discussed in Section 5.3.7 and item 35 above. For example, if the two domains identified different content or different mail servers, it would be possible for part of the set to be taken over by an attacker without damage to the other set. Consolidating the

information so the two sets could not be distinguished would, in that case, be forensically disasterous.

37. Section 6
As in earlier sections, the introductory material to this section does not pay sufficient attention to the multiplier effects associated with FQDNs with delegated variants at several levels.

In paragraph 4, a statement appears that "…end users expect consistent support of variants across all levels of the DNS tree.".   This should be qualified by two things, the first of which is the discussion of user learning in item 20 above. Second, it is important to understand that consistency with user expectations is always going to be subjective and per-user: what one user may expect to match or be equivalent may differ from what another user expects, at least in the absence of education about rules that are so clear that I believe they are likely to be impossible. There is one, and only one, model that can be explained to users as fully consistent and that is activation of a maximum of one variant per bundle, with no exceptions. As soon as multiple activations (and delegations) are permitted, users are going to base their expectations on extrapolations from one situation to another. It is unlikely that they will get those extrapolations right and that will lead to a perception of inconsistency.

38. Section 6.1.2
The first paragraph of this section is not obviously true as far as the client and/or server sides are concerned, at least without a great deal more specificity. See the discussion at the beginning of item 6 above.

39. Section 6.1.3
In the first bullet 5, it is not clear to me what a "Language-specific code point" is. Certainly there are code points in Unicode that are identified, within a script, as used only by one or two languages, but they pose no inherent marginal danger and are often quite significant for the relevant languages.

40. Section 6.1.4 and 6.2.6
One of the challenges to the variant model that was identified as early as the work that led to RFC 3743 and attempts to apply it was the question of what would happen if a bundle was established, names delegated, policies established, and next-level names put into use on that basis, and then the bundle was broken (or a name removed from it and assigned elsewhere) by some higher authority such as an order from a competent court. It seems to me that issue, and even some planning about what would happen if it occurred, should be part of Section 6.1.4 (or possibly some other section) and that the issue should be addressed in the context of the Section 6.2.6 requirement for the registry to ensure that variants belong to the same registrant.

41. Section 6.1.7
Despite the comments about training materials and user education above, I am concerned that this section calls for a major expansion in ICANN's scope and mission, something that should not be taken on simply on the basis of a few paragraphs on page 45 of a 56 page report.

42. Section 6.4.1

This is one of the issues that is necessitated by IDNs with or without variants, but, if the community is serious about IDNs, especially at the TLD and SLD levels, it is no longer acceptable for generally-available and heavily used basic DNS management and diagnostic tools to be IDNA-unaware and to operate only in terms of A-labels as a result. Such tools are much less complicated than the other ones listed in this section and include, e.g., basic command line DNS name query and reporting tools.   I believe a recommendation should be added to this section for ICANN to facilitate the development or upgrading of those tools.

43. Section 6.4.3

This section is confusing to me, largely because I'm not sure that some of the comments make sense. For example, it is not clear what it means for one URL (variant of not) to "resolve back to" another one — no such operation appears in the URL or URI specifications. Similarly, I don't understand how, in bullet 1, a search engine is expected to treat "variants as equivalent to primary domains", at least without the sorts of databases contemplated in item 6 above. For bullet 3, one of the arguments for variants (especially for Arabic) is precisely because some keyboards cannot handle some strings. Consequently it is not clear to me what that statement means either.

Almost uniquely in this otherwise well-researched report, I think this section needs a careful technical review followed by a complete rewrite.

Thanks for your time and consideration.